

Conference Abstract

Rapid Creation of a Data Product for the World's Specimens of Horseshoe Bats and Relatives, a Known Reservoir for Coronaviruses

Erica Krimmel[†], Austin Mast[†], Deborah Paul[†], Robert Bruhn[†], Nelson Rios[§], David Peter Shorthouse[|]

[†] Florida State University, Tallahassee, United States of America

[§] Yale University, New Haven, United States of America

[|] Agriculture & Agri-Food Canada, Ottawa, Canada

Corresponding author: Erica Krimmel (ekrimmel@fsu.edu)

Received: 28 Sep 2020 | Published: 02 Oct 2020

Citation: Krimmel E, Mast A, Paul D, Bruhn R, Rios N, Shorthouse DP (2020) Rapid Creation of a Data Product for the World's Specimens of Horseshoe Bats and Relatives, a Known Reservoir for Coronaviruses. Biodiversity Information Science and Standards 4: e59067. <https://doi.org/10.3897/biss.4.59067>

Abstract

Genomic evidence suggests that the causative virus of COVID-19 (SARS-CoV-2) was introduced to humans from horseshoe bats (family Rhinolophidae) (Andersen et al. 2020) and that species in this family as well as in the closely related Hipposideridae and Rhinonycteridae families are reservoirs of several SARS-like coronaviruses (Gouilh et al. 2011). Specimens collected over the past 400 years and curated by natural history collections around the world provide an essential reference as we work to understand the distributions, life histories, and evolutionary relationships of these bats and their viruses. While the importance of biodiversity specimens to emerging infectious disease research is clear, *empowering disease researchers with specimen data is a relatively new goal for the collections community* (DiEuliis et al. 2016). Recognizing this, a team from Florida State University is collaborating with partners at [GEOLocate](#), [Bionomia](#), University of Florida, the American Museum of Natural History, and Arizona State University to produce a deduplicated, georeferenced, vetted, and versioned data product of the world's specimens of horseshoe bats and relatives for researchers studying COVID-19. The project will serve as a model for future rapid data product deployments about biodiversity specimens.

The project underscores the value of biodiversity data aggregators [iDigBio](#) and the [Global Biodiversity Information Facility \(GBIF\)](#), which are sources for 58,617 and 79,862 records, respectively, as of July 2020, of horseshoe bat and relative specimens held by over one hundred natural history collections. Although much of the specimen-based biodiversity data served by iDigBio and GBIF is high quality, it can be considered raw data and therefore often requires additional wrangling, standardizing, and enhancement to be fit for specific applications. The project will create efficiencies for the coronavirus research community by producing an enhanced, research-ready data product, which will be versioned and published through Zenodo, an open-access repository (see doi.org/10.5281/zenodo.3974999).

In this talk, we highlight lessons learned from the initial phases of the project, including deduplicating specimen records, standardizing country information, and enhancing taxonomic information. We also report on our progress to date, related to enhancing information about agents (e.g., collectors or determiners) associated with these specimens, and to georeferencing specimen localities. We seek also to explore how much we can use the added agent information (i.e., [ORCID IDs](#) and [Wikidata Q identifiers](#)) to inform our georeferencing efforts and to support crediting those collecting and doing identifications. The project will georeference approximately one third of our specimen records, based on those lacking geospatial coordinates but containing textual locality descriptions.

We furthermore provide an overview of our holistic approach to enhancing specimen records, which we hope will maximize the value of the bat specimens at the center of what has been recently termed the "extended specimen network" (Lendemer et al. 2020). The centrality of the physical specimen in the network reinforces the importance of archived materials for reproducible research. Recognizing this, we view the collections providing data to iDigBio and GBIF as essential partners, as we expect that they will be responsible for the long-term management of enhanced data associated with the physical specimens they curate. We hope that this project can provide a model for better facilitating the reintegration of enhanced data back into local specimen data management systems.

Keywords

natural history collection, COVID-19, biodiversity informatics

Presenting author

Erica Krimmel

Presented at

TDWG 2020

Acknowledgements

We would like to acknowledge our partners Pam Soltis at the Univ. of Florida, Nancy Simmons at the American Museum of Natural History, and Nathan Upham at Arizona State Univ., as well as all of the collections professionals involved in curating and digitizing specimens included in our data.

Funding program

National Science Foundation (NSF) Program: COVID-19 Research

Grant title

NSF DBI 2033973: Rapid Creation of a Data Product for the World's Specimens of Horseshoe Bats and Relatives, a Known Reservoir for Coronaviruses

References

- Andersen K, Rambaut A, Lipkin WI, Holmes E, Garry R (2020) The proximal origin of SARS-CoV-2. *Nature Medicine* 26 (4): 450-452. <https://doi.org/10.1038/s41591-020-0820-9>
- DiEuliis D, Johnson K, Morse S, Schindel D (2016) Opinion: Specimen collections should have a much bigger role in infectious disease research and response. *Proceedings of the National Academy of Sciences* 113 (1): 4-7. <https://doi.org/10.1073/pnas.1522680112>
- Gouilh MA, Puechmaille S, Gonzalez J, Teeling E, Kittayapong P, Manuguerra J (2011) SARS-Coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. *Infection, Genetics and Evolution* 11 (7): 1690-1702. <https://doi.org/10.1016/j.meegid.2011.06.021>
- Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER, Bentley A, LeVan K, Bates J, Jennings D, Contreras D, Lagomarsino L, Mabee P, Ford LS, Guralnick R, Gropp RE, Revelez M, Cobb N, Seltmann K, Aime MC (2020) The Extended Specimen Network: A Strategy to Enhance US Biodiversity Collections, Promote Research and Education. *BioScience* 70 (1): 23-30. <https://doi.org/10.1093/biosci/biz140>